ADA114476

MRC Technical Summary Report #2323

CUBIC LACK OF FIT FOR THREE-LEVEL SECOND
ORDER RESPONSE SURFACE DESIGNS

N. R. Draper

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

January 1982

Received December 1, 1981

DTIC
ELECTE
MAY 18 1982
S
E
D

DTIC FILE COPY

82   05   18   048

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

CUBIC LACK OF FIT FOR THREE-LEVEL

SECOND ORDER RESPONSE SURFACE DESIGNS[†]

N. R. Draper

Technical Summary Report #2323

January 1982

ABSTRACT

A recent paper by Box and Draper (1982) discussed the detection of cubic lack of fit in second order composite design experiments, and its possible removal by the use of power transformations in the predictor variables. The designs examined were five-level designs whose coded predictor variables could assume levels $(-\alpha, -1, 0, 1, \alpha)$ for $\alpha \neq 1$ (and, typically, $\alpha > 1$). When $\alpha = 1$, only three levels exist in the design and certain singularities occur. Cubic interaction contrasts exist, but it becomes impossible to estimate the power transformations, as previously when $\alpha \neq 1$. This note describes how this happens.

AMS(MOS) Subject classification: 62J02, 62J05, 62K99

Key words: Lack of fit, Second order designs, Transformations on predictors

Work unit 4 - Statistics and Probability

CUBIC LACK OF FIT FOR THREE-LEVEL
SECOND ORDER RESPONSE SURFACE DESIGNS[†]

N. R. Draper

1. INTRODUCTION

We refer the reader to Box and Draper (1982) for notation and intro-
ductory material. The case where a first order model is fitted is unchanged,
and we consider only the second order case in what follows.

2 EXAMPLE, k = 2.

Suppose we have a two-factor composite design consisting of a replicated
"cube" (a square for $k = 2$), a "star" with axial distance $\alpha = 1$, and six center
points. Table 1 shows such a design, with data manufactured exactly as in Box
and Draper (1982, Section 3 and Appendix A). In fact, this example is an
adaption of the previous example in which $\alpha = 2$, and only the four star points
have altered response values. Two blocks are defined, as before, but the
blocks column is now not orthogonal to the x_1^2 and x_2^2 columns. The fitted
least squares equation is

$$\hat{y} = 30.11 - 5.14x_1 - 8.14x_2 - 2.22x_1^2 + 0.28x_2^2 - 3.41x_1x_2$$

$$\pm 0.35 \pm 0.21 \quad \pm 0.21 \quad \pm 0.44 \quad \pm 0.44 \quad \pm 0.24$$

where the second order coefficients and constant term have been estimated
with blocks in the model, but the term -0.09 (blocks), with coefficient s.e.
0.24, has been omitted, and where $\pm$ limits beneath each estimated coefficients

Table 1. A three-level composite design for k = 2 predictor variables and its associated estimator columns; $n_c = 8$, $n_{co} = 1$, $n_s = 5$, $\alpha = 1$.

I	x_1	x_2	x_1^2	x_2^2	$x_1 x_2$	$-5x_{122}$	$-5x_{112}$	CC	Blocks	y
1	-1	-1	1	1	1	1	1	1	-1	37.5
1	-1	-1	1	1	.1	1	1	1	-1	38.3
1	1	-1	1	1	-1	-1	1	1	-1	34.7
1	1	-1	1	1	-1	-1	1	1	-1	35.1
1	-1	1	1	1	-1	1	-1	1	-1	27.7
1	-1	1	1	1	-1	1	-1	1	-1	29.2
1	1	1	1	1	1	-1	-1	1	-1	12.2
1	1	1	1	1	1	-1	-1	1	-1	11.4
1	.	.	.	.	.	.	.	-8	-1	30.1
1	-1	.	1	.	.	-4	.	-4	1	33.8
1	1	.	1	.	.	4	.	-4	1	21.7
1	.	-1	.	1	.	.	-4	-4	1	38.4
1	.	1	.	1	.	.	4	-4	1	22.1
1	.	.	.	.	.	.	.	3.2	1	30.9
1	.	.	.	.	.	.	.	3.2	1	29.9
1	.	.	.	.	.	.	.	3.2	1	29.3
1	.	.	.	.	.	.	.	3.2	1	29.6
1	.	.	.	.	.	.	.	3.2	1	30.6

Note: The CC contrast compares the average responses at the central and the noncentral points in each block.

indicate estimated standard errors, using the pure error estimate $s_e^2 = 0.457$ to estimate σ^2. An associated analysis of variance table is shown as Table 2.

When $\alpha \neq 1$, the four third order terms pair up as $(x_1^3, x_1 x_2^2)$, $(x_2^3, x_1^2 x_2)$ and, when these pairs are orthogonalized to $\underset{\sim}{X}$ vectors of lower order, yield two vectors (per pair) which are multiples of one another. We thus obtain two "third order vectors" previously called x_{111} and x_{222} which provide cubic orthogonal contrasts in the x_1 and x_2 directions.

When $\alpha = 1$, however, $x_i^3 = x_i$ throughout. We are thus left only with the cubic vectors $[x_1 x_2^2]$ and $[x_1^2 x_2]$. These are orthogonalized against all $\underset{\sim}{X}$ vectors of lower order to give vectors $[x_{122}]$ and $[x_{112}]$ shown (without square brackets and amplified by a convenient factor of -5) in Table 1. This is why the lack of fit in Table 2 is attributed to b_{122} and b_{112}, and not b_{111} and b_{222} as previously when $\alpha \neq 1$.

We see from Table 2 that the lack of fit tests are all nonsignificant, in spite of the fact that the surface is the same non-quadratic one as we had before, for which cubic lack of fit was apparent. The pulling in of the star points from levels ± 2 to levels ± 1 has rendered pure cubic non-quadraticity undetectable, in fact. Why this is so is apparent from the geometry. The geometrical meaning of the cubic contrast illustrated in Figure 1 of the previous paper no longer applies. Whereas before we were examining the difference between the slopes of two chords whose slopes would be equal if the model were actually quadratic in the x_i dimension, we are now contrasting two unequally weighted estimates of the slope of the same chord. Thus while cubic interaction coefficients can still cause distortion, as is clear from the expected value in Eq. (3.3), setting $\alpha = 1$ has lost us our ability to detect the presence of β_{iii}.

Table 2. Analysis of variance associated with the second order model and its checks, for the data of Table 1.

Source	df		SS		MS		F
Mean	1		15,167.014		-		-
Blocks	1		5.667		5.667		10.83
First order	2		926.792		463.396		885.88
Second order extra	3		104.882		34.961		66.84
Lack $\lbrace b_{122}$	3 $\lbrace$ 1	2.097 $\lbrace$	2.070	0.699 $\lbrace$	2.070	1.53 $\lbrace$	4.53
of $\lbrace b_{112}$	1		0.000		0.000		0.00
fit $\lbrace$ CC	1		0.027		0.027		0.06
Pure error	8		3.657		0.457		
Total	18		16,210.110				

Notes: (a) Roundoff error of 0.001 occurs in the addition.
(b) First order is orthogonal to everything else.
(c) Mean, blocks, and pure quadratic are not mutually orthogonal. Thus the breakup (in more detail) is mean, blocks, and second order given mean and blocks. "Blocks" is significant only if taken out in the order indicated.
(d) Lack of fit tests provide non-significant results at the $\alpha = 0.05$ level. Tests for regression thus use $s^2 = (2.097+3.657)/11 = 0.523$ as denominator.

3. GENERAL FORMULAS

In general, a <u>three-level</u> composite design contains

(a) A "cube", consisting of a 2^k factorial, or a 2^{k-p} fractional factorial, made up of points of the type $(\pm 1, \pm 1, \ldots, \pm 1)$, of resolution $R \geq 5$ (Box and Hunter, 1961) replicated $f(\geq 1)$ times. There are thus $n_c = f2^{k-p}$ such points (where p may be zero).

(b) A "star", that is, $2k$ points $(\pm 1, 0, 0, \ldots, 0)$, $(0, \pm 1, 0, \ldots, 0), \ldots,$ $(0, 0, 0, \ldots, \pm 1)$ on the predictor variable axes, replicated r times, so that there are $n_s = 2kr$ points in all.

(c) Center points $(0, 0, \ldots, 0)$, n_0 in number, of which n_{co} are in cube blocks and n_{so} in star blocks.

It can be shown that, for any such design, k sets of columns can be isolated with the ith set containing the k columns $x_i x_j^2$, $i \neq j = 1, 2, \ldots, k$. This ith set is associated with a single vector $x_{\sim ijj}$ which is orthogonal to the $(k+1)(k+2)/2$ columns required for fitting the second degree equation and is also orthogonal to the $(k-1)$ similarly constructed vectors $x_{\ell jj}$, $\ell \neq i$. The x_i^3 vectors do not feature in these sets of columns as they do when $\alpha \neq 1$, because $x_i^3 = x_i$ throughout, now.

The elements of these vectors are such that:

for the cube points, $x_{ijj} = \phi x_i$, with $\phi = 2r/(n_c + 2r)$,

for the star points, $x_{ijj} = \gamma x_i$, with $\gamma = -n_c/(n_c + 2r)$,

for the center points $x_{ijj} = 0 = x_i$.

Thus, the k estimates of third order lack of fit, c_{31}, c_{32},..., c_{3k} are

$$c_{3i} = -\left\{\frac{\bar{y}_{\alpha i} - \bar{y}_{-\alpha i}}{2} - \frac{\bar{y}_{1i} - \bar{y}_{-1i}}{2}\right\} \quad (3.1)$$

with standard deviation

$$\sigma_{c_3} = \left\{\frac{1}{n_c} + \frac{1}{2r}\right\}^{1/2} \sigma . \quad (3.2)$$

Also

$$E(c_{3i}) = \sum_{j\neq i}^{k} \beta_{ijj} \quad (3.3)$$

and the contribution to the lack of fit sum of squares is

$$SS(c_{3i}) = c_{3i}^2 / \left\{\frac{1}{n_c} + \frac{1}{2_r}\right\} . \quad (3.4)$$

The result of Eq. (3.3) implies that effects due to β_{iii} are no longer detectable via c_{3i} when $\alpha = 1$. (Compare Eq. (3.3) with Eq. (3.7) of the earlier paper.)

4. INABILITY TO ESTIMATE TRANSFORMATION PARAMETERS

When $\alpha = 1$, it is impossible to carry through the procedure to estimate λ_i for the transformation $\xi_i^{\lambda_i}$ of the predictor variables. Consider Eqs. (3.12), (3.13), and (3.14) of the previous paper. A composite design in general ($\alpha \neq 1$) does not enable us to estimate all cubic terms, and when $\alpha = 1$, we specifically cannot estimate the β_{iii}. Thus the previous Eq. (3.12) must be replaced by

$$\beta_{ijj}x_1 x_j^2 + \beta_{2\ell\ell}x_2 x_\ell^2 + \ldots + \beta_{kmm}x_k x_m^2 \tag{4.1}$$

where $j, \ell, \ldots, m$ can take any specific values except (respectively), $1, 2, \ldots, k$. Eq. (4.1) represents the third order terms whose parameters can be estimated using a composite design with $\alpha = 1$. If the actual response could be represented by a full third order model, the estimates b_i and b_{ijj} obtained from the composite design with $\alpha = 1$ would have expectations

$$E(b_i) = \eta_i + \frac{1}{6}\eta_{iii}, \tag{4.2}$$

$$E(b_{ijj}) = \frac{1}{2}\sum_{j \neq i}^{k} \eta_{ijj}, \tag{4.3}$$

these equations replacing (3.13) and (3.14) of the original paper. Now, if a second order representation in transformed predictor variables $\xi_i^{\lambda_i}$ were indeed satisfactory, all third derivatives of the third order model with respect to the $\xi_i^{\lambda_i}$ would be zero. Requiring this leads to the conditions (where the δ_i are known constants)

$$n_{ij\ell} = 0, \quad \text{all } i \neq j = \ell = 1,2,\ldots,k; \tag{4.4}$$

$$n_{ijj} + \delta_j(1-\lambda_j)n_{ij} = 0, \qquad i \neq j = 1,2,\ldots,k; \tag{4.5}$$

$$n_{iii} + 3\delta_i(1-\lambda_i)n_{ii} + \delta_i^2(1-\lambda_i)(1-2\lambda_i)n_i = 0, \quad i = 1,2,\ldots,k. \tag{4.6}$$

Eq. (4.4) provides the (otherwise obvious) conclusion that the possibility of second order representation in the transformed variables is contra-indicated if one or more interaction estimates $b_{ij\ell}$ are non-zero.

We would now like to use b_i and b_{ijj} as estimates of their expectations in (4.2) and (4.3) and substitute these into (4.5) and (4.6) to give equations for the λ_i. (We would also need to substitute estimates for n_{ij} and n_{ii} which involve b_{ii} and b_{ij} but these are not involved in the argument which follows.) It soon becomes apparent that the combination of n_i and n_{iii} in Eq. (4.2) and the lack of either in Eq. (4.3) makes it impossible to manipulate Eqs. (4.5) and (4.6) into a suitable set of equations for the λ_i, as was achieved in the $\alpha \neq 1$ case.

The moral is simply that the choice of only three levels of the predictor variables is inadequate to the task at hand. When $\alpha = 1$, pure cubic lack of fit is not detectable and the transformation parameters λ_i cannot be estimated.

(If an attempt is made via nonlinear estimation methods to fit a second order model in $\xi_i^{\lambda_i}$ to the data, a highly ill-conditioned sum of squares surface results with consequent problems of convergence. The correlation matrix for the estimates has several entries close to 1, indicating severe overparameterization, as would be expected.)

SUMMARY

An earlier paper by Box and Draper (1982) discussed lack of fit for certain first and second order designs and their ability to detect certain kinds of lack of fit of degree one higher than has been fitted. Composite designs with five levels were explored for the second order situation. In this note, we discuss three-level composite designs and show that, while they allow the examination of certain types of third order curvature, they do not permit estimation of power transformations in the predictor variables to achieve second order representation in the transformed variables. The moral is that, <u>for this sort of transformation estimation</u>, five levels are essential.

ACKNOWLEDGEMENTS

REFERENCES

Box, G. E. P. and Draper, N. R. (1982). Measures of lack of fit for response surface designs and predictor variable transformations. <u>Technometrics</u>, 23, 000-000.

Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} fractional factorial designs, Part I. <u>Technometrics</u>, 3, 311-351; Part II, 449-458.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2323	2. GOVT ACCESSION NO. AD-A114476	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CUBIC LACK OF FIT FOR THREE-LEVEL SECOND ORDER RESPONSE SURFACE DESIGNS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) N. R. Draper		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041 DAAG29-80-C-0113
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of 610 Walnut Street Wisconsin Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 4-Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE January 1982
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Lack of fit, Second order designs, Transformations on predictors

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A recent paper by Box and Draper (1982) discussed the detection of cubic lack of fit in second order composite design experiments, and its possible removal by the use of power transformations in the predictor variables. The designs examined were five-level designs whose coded predictor variables could assume levels ($-\alpha$, $-1,0,1,\alpha$) for $\alpha \neq 1$ (and, typically, $\alpha>1$). When $\alpha=1$, only three levels exist in the design and certain singularities occur. Cubic interaction contrasts exist, but it becomes impossible to estimate the power transformations, as previously when $\alpha \neq 1$. This note described how this happens.